

Forecasting Mixed Frequency Time Series with ECM-MIDAS Models

Citation for published version (APA):

Hecq, A. W., Götz, T. B., & Urbain, J. R. Y. J. (2012). *Forecasting Mixed Frequency Time Series with ECM-MIDAS Models*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 012 <https://doi.org/10.26481/umamet.2012012>

Document status and date:

Published: 01/01/2012

DOI:

[10.26481/umamet.2012012](https://doi.org/10.26481/umamet.2012012)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Thomas B. Götz, Alain Hecq, Jean-
Pierre Urbain

**Forecasting Mixed Frequency
Time Series with ECM-MIDAS
Models**

RM/12/012

METEOR

Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

FORECASTING MIXED FREQUENCY TIME SERIES WITH ECM-MIDAS MODELS

Thomas B. Götz* Alain Hecq
Jean-Pierre Urbain

Maastricht University, SBE, Department of Quantitative Economics

March 6, 2012

Abstract

This paper proposes a mixed-frequency error-correction model in order to develop a regression approach for non-stationary variables sampled at different frequencies that are possibly cointegrated. We show that, at the model representation level, the choice of the timing between the low-frequency dependent and the high-frequency explanatory variables to be included in the long-run has an impact on the remaining dynamics and on the forecasting properties. Then, we compare in a set of Monte Carlo experiments the forecasting performances of the low-frequency aggregated model and several mixed-frequency regressions. In particular, we look at both the unrestricted mixed-frequency model and at a more parsimonious MIDAS regression. Whilst the existing literature has only investigated the potential improvements of the MIDAS framework for stationary time series, our study emphasizes the need to include the relevant cointegrating vectors in the non-stationary case. Furthermore, it is illustrated that the exact timing of the long-run relationship does not matter as long as the short-run dynamics are adapted according to the composition of the disequilibrium error. Finally, the unrestricted model is shown to suffer from parameter proliferation for small sample sizes whereas MIDAS forecasts are robust to over-parameterization. Hence, the data-driven, low-dimensional and flexible weighting structure makes MIDAS a robust and parsimonious method to follow when the true underlying DGP is unknown while still exploiting information present in the high-frequency. An empirical application illustrates the theoretical and the Monte Carlo results.

JEL Codes: C22, C53

Keywords: ECM, MIDAS, Forecasting

*Correspondence to: Thomas B. Götz, Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: t.gotz@maastrichtuniversity.nl, Tel.: +31 43 388 3578, Fax: +31 43 388 20 00

We thank J. Isaac Miller very much for useful comments on the first draft of the paper.

1 Introduction

In economics we are often concerned with the problem of dealing with variables that are available, and thus sampled, at different frequencies. One example is the task of forecasting or nowcasting the quarterly gross domestic product using monthly variables such as the industrial production index or daily indicators such as interest rates or stock prices. The classical way to deal with such a situation is the temporal aggregation of the high-frequency variables, i.e. to sample at the common low-frequency. However, this approach might lead to a loss of information due to omitting the high-frequency observations (Andreou et al., 2010). Hence, forecasting performances might improve by making use of these extra information.

In many cases, including all potential unrestricted high-frequency explanatory variables is unfeasible with regard to the number of observations for the low-frequency dependent variable. This has motivated the introduction of (Mi)xed (Da)ta (S)ampling (henceforth MIDAS) which aims at transforming a high-frequency variable so that all information in the high-frequency can be preserved (Ghysels et al., 2004). One of the virtues of the MIDAS approach is its parsimony, meaning that by a clever choice of a restricted lag polynomial we can reduce the number of parameters to estimate while still employing the information present in the high-frequency variables and its lags (Ghysels and Valkanov, 2006).

In almost all the literature, MIDAS is applied to stationary time series or to transformations of non-stationary variables like first differences. For instance, Clements and Galvao (2007) and Clements and Galvao (2009) consider the growth rate of GDP or the term spread. Andreou and Kourtellis (2010) forecast US economic activity employing a large cross-section of transformed daily series.

Merely working with differenced variables might disregard a possible long-run relationship between the variables. In this paper, we instead consider mixed-frequency time series y_t and x_t that are $I(1)$ and possibly cointegrated. We compare the forecasting performance of models for Δy_t when only the first difference terms of y_t and of x_t are included, omitting the presence of a long-run relationship, with a mixed-frequency error correction model (ECM hereafter). We also compare our mixed-frequency model with a model resulting from the temporal aggregation of high-frequency variables. In the usual case, i.e. the common-frequency framework, Clements and Hendry (1998) among others have compared models in terms of their ability to predict the levels, the first differences and the long-run relationships. Concerning the first differences, the gain of including the long-run relationship is apparent only for short horizons, say from one until 5-step ahead in their simulations. In terms of forecasting and nowcasting business cycle indicators such as inflation or the unemployment rate, however, this can be important.

Seong et al. (2007) also model cointegrated multivariate time series of mixed frequencies. However, they regard the low-frequency observations as missing data and treat them with high-frequency data by employing a state-space model. Instead, this paper aims at providing tools for working in a mixed-frequency framework but without relying on a state-space form. It turns out that most macroeconomic variables of interest to be forecasted (gross domestic product, inflation, etc) are available at a lower frequency than explanatory indicators, and therefore a regression approach makes sense. To the authors' knowledge, the evaluation of a mixed-

frequency error correction model without adopting a state-space model has not been done yet and is one of the contributions of this paper.¹

The rest of the paper is organized as follows. In Section 2 we introduce a mixed-frequency ADL model and derive the short-run dynamics implied by alternative choices of the variables to be included in a long-run component. Section 3 discusses an alternative mixed-frequency error-correction model and proposes an analogous approach to the Engle-Granger two-step framework (Engle and Granger, 1987) in which we first determine the cointegrating relationship before we plug it in an ECM regression. A MIDAS framework is proposed to capture the high-frequency short-run dynamics. The model is then compared with the unrestricted approach and classical methods of temporal aggregation using Monte Carlo simulations in Section 4. The method in Section 3 as well as some invariance tests are elaborated on and investigated as well. Section 5 illustrates the theoretical and the Monte Carlo results via an empirical illustration. Section 6 concludes.

2 The MF-ADL-model and ECM Representations

Let us first introduce a mixed-frequency autoregressive distributed lag model, denoted MF-ADL(p_h, p_l, q_h, q_l), for two non-stationary I(1) time series y_t and x_t . This can be represented by the following model:

$$\begin{aligned}
y_t = & c + \alpha_{10}y_{t,m-1} + \dots + \alpha_{p_h 0}y_{t,m-p_h} + \\
& \alpha_{01}y_{t-1} + \alpha_{11}y_{t-1,m-1} + \dots + \alpha_{p_h 1}y_{t-1,m-p_h} + \\
& \vdots \\
& \alpha_{0p_l}y_{t-p_l} + \alpha_{1p_l}y_{t-p_l,m-1} + \dots + \alpha_{p_h p_l}y_{t-p_l,m-p_h} + \\
& \beta_{00}x_t + \beta_{10}x_{t,m-1} + \dots + \beta_{q_h 0}x_{t,m-q_h} + \\
& \beta_{01}x_{t-1} + \beta_{11}x_{t-1,m-1} + \dots + \beta_{q_h 1}x_{t-1,m-q_h} + \\
& \vdots \\
& \beta_{0q_l}x_{t-q_l} + \beta_{1q_l}x_{t-q_l,m-1} + \dots + \beta_{q_h q_l}x_{t-q_l,m-q_h} + \epsilon_t,
\end{aligned} \tag{1}$$

or in terms of lag polynomials by

$$\begin{aligned}
\alpha_0(L_m)y_t = & c + \alpha_1(L_m)y_{t-1} + \dots + \alpha_{p_l}(L_m)y_{t-p_l} + \\
& \beta_0(L_m)x_t + \beta_1(L_m)x_{t-1} + \dots + \beta_{q_l}(L_m)x_{t-q_l} + \epsilon_t,
\end{aligned} \tag{2}$$

where the subscript l denotes the low-frequency variables and h represents high-frequency ones. The index t represents the low frequency and runs from 1 to T . The number of high-frequency

¹Simultaneously and independent from this work, Miller introduced cointegrating MIDAS regressions and a MIDAS test (Miller, 2011a). He shows that nonlinear least squares consistently estimate the minimum mean-squared forecast error parameter vector and derives the asymptotic distribution of the difference between the estimator and this minimum. These results are robust to errors that are correlated serially or with the regressors, extending the results of Andreou et al. (2010). In yet another paper, he generalizes the work of Chambers (2003) and Pons and Sansó (2005) on efficient estimation of the cointegrating vector. In particular, allowing for general weighting schemes, conditional and unconditional efficiency bounds are derived and a canonical cointegrating regression approach is proposed since it attains the aggregation-conditional bound asymptotically (Miller, 2011b).

observations per t -period equals m (in a year/month example, $m = 12$). ϵ_t is assumed to be a martingale difference sequence with respect to the past information and for simplicity, x_t is assumed to be strongly exogenous for the long run parameters.² In the second representation, $\alpha_k(\cdot)$ and $\beta_v(\cdot)$ (for $k = 1, \dots, p_l$ and $v = 1, \dots, q_l$) denote polynomials in L_m which stands for the lag operator in the high frequency such that $L_m^j x_{t,m-i} \equiv x_{t,m-i-j}$. Note that all the elements of both sets of polynomials are of the orders p_h and q_h , respectively. This representation will prove useful in the remainder of the paper. Note also that at this stage we abstract from deterministic trends and higher order deterministic components.

Using this notation, a variable has two indices, t and $m - i$ (for $0 \leq i \leq m - 1$). If $i = 0$, the m -index is suppressed such that $x_{t,m} \equiv x_t$ meaning it is the end-of-period observation such as the last day in a month for instance. $x_{t,1}$ is the first day of the month in a month/day analysis or the first month of the year in a year/month setting. Finally, $x_{t,m-m} \equiv x_{t-1}$. The following table illustrates this notation by the help of the year/month example.

Notation	$t = 2011, m = 12$
$x_{t+1,m-(m-1)}$	$x_{2012,Jan}$
$x_t \equiv x_{t,m}$	$x_{2011,Dec}$
$x_{t,m-1}$	$x_{2011,Nov}$
$x_{t,m-2}$	$x_{2011,Oct}$
\vdots	\vdots
$x_{t,m-(m-2)}$	$x_{2011,Feb}$
$x_{t,m-(m-1)}$	$x_{2011,Jan}$
$x_{t-1} \equiv x_{t,m-m} \equiv x_{t-1,m}$	$x_{2010,Dec}$
$x_{t-1,m-1}$	$x_{2010,Nov}$

Note that throughout the paper, Δ denotes the usual difference operator in the low-frequency: $\Delta y_t = y_t - y_{t-1}$. The difference operator in the high frequency is denoted by a subscript m : $\Delta_m y_{t,m-i} = y_{t,m-i} - y_{t,m-i-1}$.

Several models are nested in the above formulation (1) and (2). These are for instance:

- MF-ADL(0, p_l, q_h, q_l) where the high frequency observations of y are not visible,
- MF-ADL(0, $p_l, 0, q_l$), i.e. the common low frequency case, where also the high frequency observations of x are left out,
- MF-ADL(0, 0, q_h, q_l) stands for the mixed frequency distributed lag model,
- MF-ADL(0, 0, $q_h, 0$) where in addition to the previous situation, no lagged low-frequency observations on x are included.

For the subsequent analysis it is important to choose what the data generating process (DGP) is. Indeed, if the DGP is the high frequency equation (1), there are some unobserved

²This assumption is made to simplify the presentation, extensions to the weakly exogenous case are direct.

variables, e.g. the high frequency y when considering one of the restricted models above. In this paper we first consider a mixed-frequency DGP such as the MF-ADL(0, p_l, q_h, q_l). Note that this "observable" MF-ADL is often used as DGP in the MIDAS literature (see, for example, Andreou et al., 2010). As an invariance test, we will consider a high-frequency DGP in Section 4.4 of the paper. Furthermore, the resulting low-frequency observations of y_t are treated as end-of-period observations which is reasonable given that they usually become available at the end of the corresponding period.

Let us start for simplicity with

$$\begin{aligned} y_t &= c + \beta_0 x_t + \beta_1 x_{t,m-1} + \cdots + \beta_q x_{t,m-q} + \epsilon_t \\ &= c + \beta(L_m)x_t + \epsilon_t, \end{aligned} \tag{3}$$

the restricted MF-ADL(0, 0, q_h , 0) where no lagged observations on y_t 's or $s_{t,m-i}$'s are included. Note that this is a usual framework on which previous work on mixed-frequency data is applied (e.g. Andreou et al. (2010)). For convenience, we (temporarily) denote q_h by q , with $q \leq m-1$. As an example of the latter inequality, we can consider $m = 20$ days per month but only take the more recent 15 days, including the current period ($q = 14$ then). We assume x_t is $I(1)$ and that the series are cointegrated.

Given that we will rely on a simple two-step approach *à la* Engle and Granger (1987), we first consider the long-run relationship between y_{t-1} and some observation of x . We may consider three possible cases for the disequilibrium error z_{t-1} :

- (i) $z_{t-1} = y_{t-1} - \gamma x_{t-1}$: 'same-period'-case,
- (ii) $z_{t-1} = y_{t-1} - \gamma x_{t,m-i}$ with $i < m$: 'x-after-y'-case,
- (iii) $z_{t-1} = y_{t-1} - \gamma x_{t-1,m-i}$ with $i < m$: 'x-before-y'-case.

The first case corresponds to the standard framework in which the two series are sampled at the same moment. In the mixed frequency modeling there are alternative intuitive possibilities, however. For real-time data and given the release of the data by statistical offices, the 'x-after-y case' is appealing. For instance, imagine that in February 2012 we have the industrial production index for January 2012. The 'x-after-y'-case allows us to use daily information available in February 2012 (interest rate, stock prices,..) to test for cointegration.

By relying on the representation that employs lag polynomials, it will be illustrated that for every case an ECM representation of the following form can be constructed:

$$\Delta y_t = c - (y_{t-1} - \beta(1)L_m^\phi x_t) + \beta^*(L_m)(1 - L_m)x_t + \epsilon_t,$$

with $\phi = \{i, m, m+i\}$ and $1 \leq i \leq m-1$ such that the respective $x_{t,m-i}$ -observation enters z_{t-1} . The respective transformation applied to the lag polynomial, $\beta(L_m)$, is either similar or identical to the Beveridge-Nelson decomposition (see, for instance, Davidson, 2000). As will be shown in the sequel, the timing of the high-frequency observation appearing in the long-run relationship impacts the derivation of the short-run dynamics $\beta^*(L_m)$ that might be captured

in a MIDAS framework. Although the timing will not affect the asymptotic properties of tests statistics, the Monte Carlo section evaluates the impact of that choice for small samples.

Note that at this stage we know the underlying DGP such that we are able to determine, at the model representation level, the structure of the short-run dynamics coherent with any timing under consideration. In practice, the DGP as well as the correct timing and therefore the according amount of short-run variables to include may be unknown. Using the Monte Carlo results in Section 4 we will argue that over-parameterizing the short-run dynamics will solve the above mentioned problem for a practitioner. However, the implications of different timings in the long-run term on the short-run dynamics at the model representation level are of importance for understanding the Monte Carlo study (and its results).

2.1 'Same-period'

If z_{t-1} consists of y_t - and $x_{t,m-i}$ -observations of the same period, the ECM-representation can easily be derived by noting that the lag polynomial, $\beta(L_m) = \beta_0 x_t + \beta_1 x_{t,m-1} + \dots + \beta_q x_{t,m-q}$ can be rewritten in the following way:

$$\begin{aligned}\beta(L_m) &= (\sum_{j=0}^q \beta_j) L_m^m + \left[\beta_0 + (\beta_0 + \beta_1) L_m + \dots + (\sum_{j=0}^q \beta_j) L_m^q + \dots + (\sum_{j=0}^q \beta_j) L_m^{m-1} \right] (1 - L_m) \\ &= \beta(1) L_m^m + \beta^*(L_m) (1 - L_m),\end{aligned}$$

such that $\beta^*(L_m)$ is a lag polynomial of order $m - 1$. This implies that

$$\begin{aligned}y_t &= c + \beta(L_m) x_t + \epsilon_t \\ \Leftrightarrow \Delta y_t &= c - (y_{t-1} - \beta(1) L_m^m x_t) + \beta^*(L_m) (1 - L_m) x_t + \epsilon_t \\ &= c - (y_{t-1} - \beta(1) x_{t-1}) + \beta^*(L_m) \Delta_m x_t + \epsilon_t.\end{aligned}\tag{4}$$

Note that the error correction (EC) term consists of one y_t - and one $x_{t,m-i}$ -variable only. The short-run, however, are the high-frequency differences of all the $x_{t,m-i}$ -observations between t and $t - 1$ leading to m high-frequency terms entering the short-run dynamics in total.

2.2 'x-after-y'

If the deviation from the long-run is best represented by $z_{t-1} = y_{t-1} - \gamma x_{t,m-i}$ with $i < m$, the ECM-representation takes the form

$$\begin{aligned}\Delta y_t &= c - (y_{t-1} - \beta(1) L_m^i x_t) + \beta^*(L_m) (1 - L_m) x_t + \epsilon_t \\ &= c - (y_{t-1} - \beta(1) x_{t,m-i}) + \beta^*(L_m) \Delta_m x_t + \epsilon_t,\end{aligned}\tag{5}$$

where $1 \leq i \leq m - 1$ and

$$\beta^*(L_m) = \begin{cases} \beta_0 + \dots + (\sum_{j=0}^{i-1} \beta_j) L_m^{i-1} - (\sum_{j=i+1}^q \beta_j) L_m^i - (\sum_{j=i+2}^q \beta_j) L_m^{i+1} - \dots - (\beta_q) L_m^{q-1} & \text{if } q > i \\ \beta_0 + (\beta_0 + \beta_1) L_m + \dots + (\sum_{j=0}^q \beta_j) L_m^q + \dots + (\sum_{j=0}^q \beta_j) L_m^{i-1} & \text{if } q \leq i. \end{cases}$$

Hence, in the 'x-after-y'-case, we have $\max(q, i)$ short-run dynamics terms (since $\beta^*(L_m)$ is a polynomial of order $\max(q - 1, i - 1)$).

2.3 'x-before-y'

If $z_{t-1} = y_{t-1} - \gamma x_{t-1, m-i}$ with $i < m$ is the most suitable long-run relationship, we obtain the following ECM-representation analogously to the previous cases:

$$\begin{aligned}\Delta y_t &= c - (y_{t-1} - \beta(1)L_m^{m+i}x_t) + \beta^*(L_m)(1 - L_m)x_t + \epsilon_t \\ &= c - (y_{t-1} - \beta(1)x_{t-1, m-i}) + \beta^*(L_m)\Delta_m x_t + \epsilon_t,\end{aligned}\tag{6}$$

where $L_m^{m+i}x_t = L_m^i L_m^m x_t = L_m^i x_{t-1} = x_{t-1, m-i}$ and

$$\beta^*(L_m) = \beta_0 + (\beta_0 + \beta_1)L_m + \left(\sum_{j=0}^q \beta_j\right)L_m^q + \dots \left(\sum_{j=0}^q \beta_j\right)L_m^m + \dots \left(\sum_{j=0}^q \beta_j\right)L_m^{m+i-1}.$$

Here, we deal with $m + i$ short-run dynamics terms.

2.4 Extending along q_l

If we allow for more low-frequency periods of the $x_{t, m-i}$ -variable, e.g. $q_l > 0$ in the MF-ADL-model, we obtain an ECM representation that follows from the techniques in the subsections above. Consider the MF-ADL(0, 0, q_h , q_l) model:

$$y_t = c + \beta_0(L_m)x_t + \beta_1(L_m)x_{t-1} + \dots + \beta_{q_l}(L_m)x_{t-q_l} + \epsilon_t.$$

Recalling that each of the q_l lag polynomials is of the same order (q_h), straightforward application of the techniques employed before leads to the following ECM representation:

$$\Delta y_t = c - \left[y_{t-1} - \left(\sum_{j=0}^{q_l} \beta_j(1) \right) L_m^{\phi} x_t \right] + \beta^*(L_m)(1 - L_m)x_t + \epsilon_t,\tag{7}$$

where $\beta^*(L_m)$ represents an 'aggregate' lag polynomial of order $(mq_l + q_h - 1)^3$ whose structure depends on which $x_{t, m-i}$ -variable appears in the long-run relationship which is, in turn, determined by ϕ as before.

2.5 Inclusion of an AR-term

Now, let us add an autoregressive term in the MF-ADL model. We will focus on the case with $q_l = 0$ for illustrative purposes (the case with $q_l > 0$ is trivial given the previous subsection). Hence, we consider the following MF-ADL(0, p , q , 0) model:

$$\begin{aligned}y_t &= c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta(L_m)x_t + \epsilon_t \\ \Leftrightarrow \alpha(L)y_t &= c + \beta(L_m)x_t + \epsilon_t,\end{aligned}$$

³Except in the 'x-before-y'-case with $q_l = 1$ where the order equals $(\max(q_h, i) + m - 1)$. Note that the cases when $q_l = 0$ were discussed in the sections before.

where $\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$ is a lag polynomial in the low frequency. Straightforward application of the methods employed before gives:

$$\begin{aligned} \alpha(L)y_t &= c + \beta(L_m)x_t + \epsilon_t \\ \Leftrightarrow \alpha(1)y_{t-1} + \alpha^*(L)\Delta y_t &= c + \beta(1)x_{t,m-\phi} + \beta^*(L_m)\Delta_m x_t + \epsilon_t \\ \Leftrightarrow \alpha^*(L)\Delta y_t &= c - \alpha(1)\left[y_{t-1} - \frac{\beta(1)}{\alpha(1)}x_{t,m-\phi}\right] + \beta^*(L_m)\Delta_m x_t + \epsilon_t, \end{aligned} \quad (8)$$

where $\alpha^*(L)$ is a lag polynomial of order $p-1$ (due to the application of the 'classical' Beveridge-Nelson decomposition) such that terms involving the dependent and independent variables enter the short-run dynamics.

3 The Choice of the Error-Correction Term for Cointegrating Analysis

Note that the short-run dynamics terms are high-frequency differences which can be estimated by a MIDAS regression, for example. However, the disequilibrium error consists of one particular $x_{t,m-i}$ and y_{t-1} . If we want to allow for more observations of $x_{t,m-i}$ appearing in the long-run relationship, we obtain low-frequency differences in the short-run dynamics (note that the respective difference operators do not have an m -subscript anymore):

$$\begin{aligned} y_t &= c + \beta_0 x_t + \beta_1 x_{t,m-1} + \dots + \beta_q x_{t,m-q} + u_t = c + \beta(L_m)x_t + \epsilon_t \\ \Leftrightarrow \Delta y_t &= c - (y_{t-1} - \beta_0 x_{t-1} - \beta_1 x_{t-1,m-1} - \dots - \beta_q x_{t-1,m-q}) + \\ &\quad \beta_0 \Delta x_t + \beta_1 \Delta x_{t,m-1} + \dots + \beta_q \Delta x_{t,m-q} + \epsilon_t \\ &= c - (y_{t-1} - \beta(L_m)x_{t-1}) + \beta(L_m)\Delta x_t + \epsilon_t. \end{aligned} \quad (9)$$

Focusing on this representation, one could estimate the (mixed-frequency) error-correction term in (9) via MIDAS by imposing

$$y_t = \beta(L_m)x_t = \beta \sum_{i=0}^q w_i(\theta)x_{t,m-i}, \quad (10)$$

where $w_i(\theta)$ are weights that sum up to one in order to identify the scale coefficient, β , and treat the short-run terms as common frequency variables with respect to the dependent variable. However, our approach of using only one particular observation of $x_{t,m-i}$ in the error-correcting term along with high-frequency short-run dynamics has several advantages over the above mentioned option. Indeed, testing for a MIDAS-type cointegrating relationship is computationally difficult (see Götz, 2010 for details).

Focusing on our ECM-MIDAS models, the question remains how to determine the timing of the long-run relationship, i.e. which $x_{t,m-i}$ -term enters the EC-term. Analogously to the Engle-Granger two-step framework (1987), first a certain regressor variable is determined to enter the long-run term. Second, the cointegrating relationship is estimated via fully modified least squares. In order to choose one of the $x_{t,m-i}$ -terms or a combination of them, we consider the set $x_{t+j,m-i}$, with $j \in \{0, 1\}$ and $i \in [0, m-1]$ excluding the combination $\{j=1, i=0\}$ as candidates. Hence, we screen through the possible cases described in the previous sections:

Same-period ($j = 0, i = 0$), 'x-after-y' ($j = 1, 0 < i \leq m - 1$) and 'x-before-y' ($j = 0, 0 < i \leq m - 1$). It should be clear that there are $m - 1$ 'x-before-y'- and 'x-after-y'-cases giving, with the 'same-period'-case, a total of $2m - 1$ candidates. The table below illustrates this set of $x_{t,m-i}$ -terms for a year/month example.

Notation	$t = 2010, m = 12$	Case
$x_{t,11}$	$x_{2011,Nov}$	'x-after-y'
\vdots	\vdots	
$x_{t,2}$	$x_{2011,Feb}$	'x-after-y'
$x_{t,1}$	$x_{2011,Jan}$	'x-after-y'
$x_{t-1,(12)}$	$x_{2010,Dec}$	'same-period'
$x_{t-1,11}$	$x_{2010,Nov}$	'x-before-y'
\vdots	\vdots	
$x_{t,2}$	$x_{2010,Feb}$	'x-before-y'
$x_{t,1}$	$x_{2010,Jan}$	'x-before-y'

The question whether a certain $x_{t,m-i}$ -term (and if yes, which one) or a combination of all of them (and if yes, how to combine) should enter the EC-term will be answered in the next section.

After the long-run relationship is determined and estimated as described above, we are able to estimate all remaining parameters in the corresponding ECM representation exploiting the superconsistency results of the least squares estimator of cointegrating vectors (Miller, 2011a). Due to the fact that the short-run dynamics are high-frequency variables, different candidate estimation methods are available. Their forecasting performances with or without the inclusion of the disequilibrium error term are investigated in a Monte Carlo study in the next section.

4 Monte Carlo Simulations

4.1 Data Generation Process

As already mentioned in Section 2, we first assume that the DGP is the "observable" MF-ADL($0, p_l, q_h, q_l$) that is often employed in the MIDAS-literature (see, e.g., Andreou et al., 2010). In particular, we assume that $x_{t,m-i}$ is generated as a simple random walk in the high frequency:

$$x_{t,m} = x_{t,m-1} + \epsilon_{t,m},$$

where $x_{t,1} = x_{t-1,m} + \epsilon_{t,1}$, $x_0 = u_1$ with $\epsilon_{t,m} \sim \text{i.i.d. } N(0, 1)$ and $u_1 \sim N(0, 1)$.

Before generating Δy_t using one of the ECM representations from the previous section, we have to make several assumptions. Firstly, let us assume the "best" long-run relationship is the 'same-period' case, e.g. $z_{t-1} = y_{t-1} - \gamma x_{t-1,m}$. Furthermore, we assume that the cointegrating relationship is known exactly, e.g. γ is known. Without loss of generality we set $\gamma = 1$. This is, of course, a strong assumption which should be relaxed in further research. For now, however,

we want to focus on the forecasting performance of our mixed-frequency ECMs employing different methods to estimate the short-run dynamics. Note, however, that we could also use any consistent estimator of the long-run relationship such as fully modified OLS. Finally, assume $m = 12$ (think of months in a year or weeks in a quarter), $q_h = m - 1$ and $q_l = 0$ in the DGP.

We fix $y_0 \sim N(0, 1)$ to compute $z_0 = y_0 - x_{0,m}$. Finally, y_t is generated as

$$\Delta y_1 = c + \delta z_0 + \alpha y_0 + \sum_{i=0}^{m-1} \beta_i \Delta_m x_{1,m-i} + v_1,$$

(from which $y_1 = y_0 + \Delta y_1$ to generate)

$$\Delta y_t = c + \delta z_{t-1} + \alpha \Delta y_{t-1} + \sum_{i=0}^{m-1} \beta_i \Delta_m x_{t,m-i} + v_t,$$

with $v_t \sim \text{i.i.d. } N(0, 1)$ and $y_t = y_0 + \sum_{i=1}^t \Delta y_i$ for $t > 1$. This ensures that we obtain $I(1)$ variables y_t and $x_{t,m-i}$ where the former is a low frequency and the latter is a high-frequency variable. Furthermore, y_t and the m^{th} observation of $x_{t,m-i}$ are cointegrated of order $CI(1, 1)$.

Throughout the simulations we fix $c = 0.1$, $\alpha = 0.5$ and consider four different values for the error correction coefficient $\delta \in \{-0.75, -0.25, -0.1, 0\}$. The coefficients of the high-frequency differences in $x_{t,m-i}$ are the cumulative ones of the respective $x_{t,m-i}$ -observations in levels. In particular, Section 2.1 illustrates that the coefficient on $\Delta_m x_{t,m-i}$ is the sum of the coefficients on the first i observations of $x_{t,m-i}$ in the underlying MF-ADL model. We assume that more recent observations of $x_{t,m-i}$ should have a larger impact on y_t implying that the β -coefficients above should be increasing, but with a decreasing "step size".⁴

4.2 Forecasting: Approaches and Robustness Tests

We compare eight different approaches in terms of their forecasting performances. These are:

- (1) Unrestricted short-run with cointegration,
- (2) Unrestricted short-run; long-run relationship excluded,
- (3) Restricted short-run (MIDAS) with cointegration,
- (4) Restricted short-run (MIDAS); long-run relationship excluded,
- (5) Point-in-Time sampling with cointegration,
- (6) Point-in-Time sampling; long-run relationship excluded,
- (7) Average Sampling with cointegration,

⁴If the coefficients are represented by a continuous function, it would be increasing and concave.

(8) Average Sampling; long-run relationship excluded.

'Unrestricted short-run' simply indicates that the high-frequency variables are not temporally aggregated nor are their coefficients estimated via a MIDAS regression; they are estimated via ordinary least squares. These two cases can be seen as cases where no restrictions are imposed on the coefficients of the $\Delta_m x$ -terms. Comparing (1) and (2) is essentially what Clements and Hendry (1998) or Engle and Yoo (1987) investigate in the standard framework.

As mentioned in the introduction, MIDAS aims at preserving information present in the high-frequency variables and estimating parameters in a parsimonious way. If we consider the different ECM representations given in Section 2, we see that estimating such models without any restrictions might be unappealing due to parameter proliferation (Andreou and Kourtellis, 2010). If y_t is measured at a quarterly and $x_{t,m-i}$ at a daily frequency, we might have over 50 parameters to estimate. In a MIDAS model we hyper-parameterize the polynomial lag structure yielding

$$\alpha^*(L)\Delta y_t = c + \delta z_{t-1} + \beta \sum_{j=0}^{q_l+1} \sum_{i=0}^{m-1} w_{i+j*m+1}(\theta) \Delta_m x_{t-j,m-i} + u_t,$$

where $w_j(\theta)$ are weights that sum up to one in order to identify the scale coefficient β .⁵ The weights are based on an underlying weight function for which different possible specifications are proposed in the literature (see, for instance, Ghysels et al., 2007). In this paper, we employ the exponential Almon Lag polynomial having its roots in the work of Almon (1965) on polynomial interpolation in distributed lag equations. The weights are given by

$$w_j(\theta) = w_j(\theta_1, \theta_2) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^k \exp(\theta_1 j + \theta_2 j^2)}.$$

As discussed in Ghysels et al. (2007), this low-dimensional lag polynomial specification is extremely flexible as it allows for the weight-determination to be completely data-driven, and it allows for various possible shapes of the weight function. Often, decaying weights are detected due to more recent observations being more important. Such a declining structure can be imposed by restricting $\theta_2 \leq 0$. According to Ghysels and Valkanov (2006), forecasts based on the exponential Almon lag polynomial dominate the ones based on other lag polynomial specifications.

Note that in the equation above, a single weight function is specified for the whole set of short-run variables (in line with Andreou and Kourtellis, 2010). Two comments have to be made here: First, the purely data-driven determination of the weight function allows us to include more short-run variables than theoretically necessary according to the ECM representations in Section 2 since the redundant ones will be assigned a zero weight. Assume, for example, that the long-run relationship is the same-period-case and $q_l = 0$. Suppose further that we include $2q_h$ variables in the MIDAS-regression although, according to Section 2.4, only q_h short-run

⁵Note that the index j runs until $q_l + 1$ in order to include enough short-term variables irrespective of the exact timing of the variables appearing in the disequilibrium term.

variables would be suffice. Nevertheless, the flexibility of MIDAS will result in zero weights for the redundant q_h variables. Second, it is also possible to assign one weight function to every low-frequency period of the short-run variables. Ghysels et al. (2007) present a generalization of MIDAS models nesting such an approach. Such an extension might be interesting to apply in our case since it allows the weights to change more abruptly. Note that the "joint" weight function will simply act like a smoothing curve of all the successive "single" weight functions. If the first m variables are assumed to be significant for forecasting whereas the next m are not, two separate weight functions (coming along with two scale parameters β above) will be more convenient to capture that difference.

The approaches (5)-(7) are standard methods used to temporally aggregate high-frequency data according to Silvestrini and Veredas (2008) or Marcellino (1999). Point-in-Time sampling selects one particular high-frequency-observation of a low-frequency-period (often the last one, e.g. December or the last trading day) as the corresponding low-frequency-observation $x_t = x_{t,m-i}$. In our case, it means that the $\Delta_m x$ -terms are replaced by Δx_t . As a benchmark, we will apply Point-in-Time sampling to the last observation ($i = 0$ above) in our simulations. Average sampling simply selects the average per low-frequency period as the respective low-frequency observation $x_t = \frac{1}{m} \sum_{i=0}^{m-1} x_{t,m-i}$.

To summarize the different ECM representations, the four different approaches correspond to the following forecasting models:

$$\begin{aligned}\widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta} z_{t-1} + \widehat{\alpha} \Delta y_{t-1} + \sum_{i=0}^{m-1} \widehat{\beta}_i \Delta_m x_{t,m-i} \text{ (Unrestricted)} \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta} z_{t-1} + \widehat{\alpha} \Delta y_{t-1} + \widehat{\beta} \sum_{i=0}^{m-1} w_i(\widehat{\theta}_1, \widehat{\theta}_2) \Delta_m x_{t,m-i} \text{ (MIDAS)} \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta} z_{t-1} + \widehat{\alpha} \Delta y_{t-1} + \widehat{\beta} \Delta x_{t,m} \text{ (Point-in-Time)} \\ \widehat{\Delta y_t} &= \widehat{c} + \widehat{\delta} z_{t-1} + \widehat{\alpha} \Delta y_{t-1} + \widehat{\beta} \Delta \frac{1}{m} \sum_{i=0}^{m-1} x_{t,m-i} \text{ (Averaging)}\end{aligned}$$

The forecasting performance of the various approaches is assessed for $T = 50, 100, 250$ and 500 low frequency observations. The estimated parameters are then used to compute 36 one-step-ahead forecasts of Δy employing the respective ECM representation. These forecasts are compared with the actual values and the Root Mean Squared Error (RMSE) is computed for every approach in each iteration. Finally, for each method, the mean of the RMSEs is obtained. Note that when the long-run component of the model is excluded, the second term in the models above is obviously dropped. In order to see whether there are RMSE differences between two models, i.e. whether the forecast accuracy of two competing models differs, two statistical tests are considered depending on whether the models are nested or non-nested. In the latter case, the classical Diebold-Mariano test (Diebold and Mariano, 1995) is employed whereas in the former case the modified Diebold-Mariano test proposed by Harvey et al. (1998) is conducted (see also Clark and McCracken (1999)). Note that parameter uncertainty is not taken into account.

At this stage, we have not been specific about which long-run relationship to employ in cases (1) and (3). For both approaches the "optimal" scenario is analyzed meaning that the correct, 'same-period', long-run is considered. However, to investigate the impact of choosing an erroneously dated error-correction term, we compute the RMSEs corresponding to one long-run

relationship of the 'x-before-y'- and one of the 'x-after-y'-type.⁶ In particular, taking the mid-period observations (middle of the year since $m = 12$), the 'x-before-y'- and 'x-after-y'-cases are represented by $z_{t-1}^{x\text{-before-y}} = y_{t-1} - x_{t-1,m-6}$ and $z_{t-1}^{x\text{-after-y}} = y_{t-1} - x_{t,m-6}$, respectively. Note that in the estimated models the number of terms entering the short-run dynamics employed must be adapted consistently with the specifications in Sections 2.2 and 2.3.

Finally, in order to assess the sensitivity of the unrestricted and the MIDAS approach to the number of short-run dynamics terms included, two further invariance tests will be applied. In particular, the RMSEs for cases (1)-(4) are computed employing first less variables than theoretically necessary, and second more than theoretically necessary (see Section 2) in the respective models. In particular, only the first $\frac{1}{2}m$ or as many as $3m$ (as opposed to m) short-run variables are included to investigate whether the inclusion of too few or too many variables distorts our forecasts. Here, only the models with correct long-run relationship are considered.

All computations are done using GAUSS10 with 1,000 replications.

4.3 Results

Tables 1 to 4 below show the outcomes for the 8 models. For cases (1) and (3) the results for erroneously dated error-correction terms are given. To be more specific, the figures are the (means of the) RMSEs of the respective method relative to the benchmark which is the 'same-period'-case with error-correction term. Hence, a value larger than 1 indicates the benchmark to be superior while a value smaller than 1 leads to the opposite conclusion. Furthermore, in order to investigate whether a certain method significantly outperforms the benchmark, the percentages in brackets below the ratios of RMSEs report the frequency with which the (modified) Diebold-Mariano tests reject the null of equal accuracy at the 5% significance level. Hence, the closer the percentage to 100% (0%), the more (less) important is the difference in the forecasting performance of the model at hand with respect to the benchmark case.

Let us first focus on the cases with a correctly chosen long-run relationship (labeled 'same-period' in the tables). Firstly, in the presence of cointegration, $\delta \neq 0$, excluding the disequilibrium error has a significant impact on the forecasting performance for all cases considered, MIDAS, unrestricted, Point-in-Time and Average sampling. Of course, this effect decreases the smaller the δ , i.e. the slower the speed of adjustment to the disequilibrium. The rejection frequency for the tests of equal forecast accuracy support the superiority of the models including a long-run term, for both the unrestricted cases as well as for the standard aggregation methods. This observation is in line with Clements and Hendry (1998).

Hence, for all four approaches, neglecting the EC-term if the variables are cointegrated has a significant negative impact on the forecast accuracy. As mentioned before, in the MIDAS literature, empirical analyzes often only consider $I(0)$ -variables that are first differenced transformed $I(1)$ -variables (see, for example, Andreou and Kourtellis, 2010). In the presence of cointegration, the inclusion of the respective relationship might yield considerable forecasting

⁶Note that for Point-in-Time and Average sampling the regressor observations entering the long-run term are always the corresponding temporally aggregated variables, i.e. $x_{t-1,m}$ and $\frac{1}{m} \sum_{i=0}^{m-1} x_{t-1,m-i}$, respectively.

MIDAS		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'x-before-y'	1.1283 (32.2%)	1.1326 (37.6%)	1.1309 (34.9%)	1.1248 (34.5%)
	'x-after-y'	1.4598 (75.8%)	1.4825 (82.3%)	1.5003 (84%)	1.4961 (84.1%)
Without Long-Run	'same-period'	2.7322 (100%)	2.7779 (100%)	2.8193 (100%)	2.8168 (100%)
		$\delta = -0.25$			
With Long-Run	'x-before-y'	1.0019 (15.9%)	1.0011 (9.2%)	1.0022 (8.1%)	1.0018 (6.5%)
	'x-after-y'	1.0028 (11.7%)	1.0097 (9.5%)	1.0116 (9.4%)	1.0112 (10.4%)
Without Long-Run	'same-period'	1.3254 (99.7%)	1.3402 (99.5%)	1.3559 (99.7%)	1.3522 (100%)
		$\delta = -0.1$			
With Long-Run	'x-before-y'	1.0100 (14.9%)	1.0036 (9.7%)	1.0030 (6.8%)	1.0031 (7.5%)
	'x-after-y'	1.0054 (17.7%)	1.0059 (12.1%)	1.0063 (10.6%)	1.0051 (9.9%)
Without Long-Run	'same-period'	1.0773 (75%)	1.0912 (79.4%)	1.1026 (81.9%)	1.1010 (81.1%)
		$\delta = 0$			
With Long-Run	'x-before-y'	1.0223 (23.6%)	1.0150 (14%)	1.0111 (9.7%)	1.0139 (11.7%)
	'x-after-y'	1.0019 (31.8%)	1.0016 (22.1%)	0.9996 (14.3%)	1.0008 (11%)
Without Long-Run	'same-period'	0.9041 (14.5%)	0.9634 (0.134)	0.9873 (7.5%)	0.9914 (8.7%)

Table 1: RMSEs relative to the 'same-period'-case with long run, the benchmark, and percentage number of rejection of tests of equal forecast accuracy at the 5% level compared to the benchmark, short-run dynamics modeled by MIDAS, correct number of variables

improvements. If the variables are not cointegrated, neglecting the long-run term results in the lowest RMSEs.

Consider now the impact of erroneously choosing the timing for the $x_{t,m-i}$ -observation entering the error-correction term. Firstly, employing the correct long-run relationship yields the lowest RMSE for both, MIDAS and the unrestricted approach. Secondly, as expected, choosing an erroneously dated timing for the EC-term is far less severe than not including an error-correction term at all. Note that the difference between the RMSEs of the 'same-period'-cases and the 'x-after-y'- or 'x-before-y'-cases is less than the difference between the RMSEs of the 'same-period'-cases and the 'Without Long-Run'-cases. Actually, the rejection frequency for the tests of equal forecast accuracy shows that the models with an erroneously dated timing often do not perform significantly worse than the benchmark case.⁷ The reason is that by transforming the short-run dynamics according to Section 2, the models resulting from an erroneous timing are not misspecified, but merely represent another representation.

We may summarize these simulation results as follows. If cointegration is present, it seems not to matter which error-correction term is included (unless $\delta = -0.75$) as long as we do

⁷The only case where the forecasting performance of the benchmark case is significantly better than when an erroneous timing is considered is that of employing MIDAS and considering the 'x-after-y'-case and $\delta = -0.75$.

Unrestricted		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'x-before-y'	1.104 (30%)	1.0369 (16.1%)	1.0139 (10.2%)	1.0052 (8.8%)
	'x-after-y'	1.0043 (19%)	1.0007 (12.8%)	1.0003 (9.8%)	1.0001 (7.3%)
Without Long-Run	'same-period'	2.7591 (100%)	2.7869 (100%)	2.8231 (100%)	2.8214 (100%)
		$\delta = -0.25$			
With Long-Run	'x-before-y'	1.1048 (30.8%)	1.0378 (16.5%)	1.0137 (9.5%)	1.0052 (7.8%)
	'x-after-y'	1.0051 (20.4%)	1.0006 (12.3%)	1.0002 (9.7%)	1.0001 (8.2%)
Without Long-Run	'same-period'	1.3314 (99.3%)	1.3413 (99.3%)	1.3567 (99.7%)	1.353 (100%)
		$\delta = -0.1$			
With Long-Run	'x-before-y'	1.1043 (31.3%)	1.038 (17%)	1.0137 (10%)	1.0052 (7.7%)
	'x-after-y'	1.0057 (20.4%)	1.0007 (13.9%)	1.0002 (8.8%)	1.0001 (7.9%)
Without Long-Run	'same-period'	1.0794 (73.5%)	1.092 (77.3%)	1.1028 (81.1%)	1.1011 (80%)
		$\delta = 0$			
With Long-Run	'x-before-y'	1.1021 (34.1%)	1.0376 (19.5%)	1.0134 (11.8%)	1.0052 (8%)
	'x-after-y'	1.0053 (29.4%)	1.0016 (19%)	1.003 (13%)	1.0001 (8.3%)
Without Long-Run	'same-period'	0.9123 (14.5%)	0.9644 (15.1%)	0.9891 (7.9%)	0.9959 (7.1%)

Table 2: RMSEs relative to the 'same-period'-case with long run, the benchmark, and percentage number of rejection of tests of equal forecast accuracy at the 5% level compared to the benchmark, short-run dynamics unrestricted, correct number of variables

Point-in-Time		$\delta = -0.75$			
	Timing / T	50	100	250	500
Without Long-Run	'same-period'	2.7453 (100%)	2.7628 (100%)	2.8055 (100%)	2.8065 (100%)
		$\delta = -0.25$			
Without Long-Run	'same-period'	1.3267 (99.7%)	1.3371 (99.3%)	1.3527 (99.7%)	1.3491 (100%)
		$\delta = -0.1$			
Without Long-Run	'same-period'	1.0762 (75.6%)	1.0902 (79%)	1.1024 (81%)	1.0999 (80.7%)
		$\delta = 0$			
Without Long-Run	'same-period'	0.9024 (13.6%)	0.9625 (13.7%)	0.9887 (7.6%)	0.9960 (7.8%)

Table 3: RMSEs relative to the 'same-period'-case with long run, the benchmark, and percentage number of rejection of tests of equal forecast accuracy at the 5% level compared to the benchmark, Point-in-Time sampling on $x_{t,m-i}$, correct number of variables

include one and adapt the short-run to the timing accordingly as explained in Section 2. Indeed, although not displayed here for convenience, if we chose an erroneous timing such as the 'x-before-y'-case and include only the first m instead of $m + 6$ (which would be consistent with Section 2) short-run variables, the forecasting performance significantly deteriorates (the means

Averaging		$\delta = -0.75$			
	Timing / T	50	100	250	500
Without Long-Run	'same-period'	1.9183 (100%)	1.9231 (100%)	1.9579 (100%)	1.9584 (100%)
		$\delta = -0.25$			
Without Long-Run	'same-period'	1.2155 (98%)	1.2233 (98.7%)	1.2391 (98.6%)	1.2361 (100%)
		$\delta = -0.1$			
Without Long-Run	'same-period'	1.0605 (68%)	1.0709 (69.2%)	1.0818 (70.7%)	1.0801 (68.6%)
		$\delta = 0$			
Without Long-Run	'same-period'	0.9074 (15.2%)	0.9662 (14.5%)	0.9889 (10.3%)	0.9962 (9.2%)

Table 4: RMSEs relative to the 'same-period'-case with long run, the benchmark, and percentage number of rejection of tests of equal forecast accuracy at the 5% level compared to the benchmark, Average sampling on $x_{t,m-i}$, correct number of variables

of the RMSEs almost double for some δ). The effect is similar to that of including too few short-run terms with the correct timing, as we will see later in this section. In the absence of cointegration, ignoring the EC term yields the best forecasting performance.

Let us now compare MIDAS to the other methods. Tables 5 to 7 report the RMSEs of the three methods, unrestricted short-run, Point-in-Time and Average sampling, relative to the RMSEs of MIDAS for the correct model, i.e. the 'same-period'-case with long-run term for $\delta \neq 0$ and without EC term for $\delta = 0$.

MIDAS vs Unrestricted		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'same-period'	1.1147	1.0515	1.0170	1.0075
		$\delta = -0.25$			
With Long-Run	'same-period'	1.1163	1.0522	1.0170	1.0076
		$\delta = -0.1$			
With Long-Run	'same-period'	1.1169	1.0522	1.0172	1.0075
		$\delta = 0$			
Without Long-Run	'same-period'	1.1160	1.0515	1.0170	1.0075

Table 5: RMSEs, Unrestricted relative to MIDAS, correct number of variables

It appears that MIDAS always performs better than leaving the short-run unrestricted or than using Average sampling. This is particularly true for small T . This illustrates the parameter proliferation problem of the unrestricted short-run method (see Andreou and Kourtellis, 2010) noting that one has to estimate 9 extra parameters compared to the case when MIDAS is used (the effect increases with m , of course). The difference is very small, however, for $T = 500$ such that the methods almost coincide asymptotically. Average sampling yields the worst results for large sample sizes. However, note that the $x_{t,m-i}$ -observation entering the long-run term is always a particular high-frequency observation as opposed to the average of all high-frequency observations per low-frequency period. Furthermore, the short-run terms in the DGP are generated with non-flat weights. Given these two points, the disappointing results

MIDAS vs Point-in-Time		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'same-period'	0.9739	0.9914	1.0014	1.0024
		$\delta = -0.25$			
With Long-Run	'same-period'	0.9746	0.9918	1.0014	1.0025
		$\delta = -0.1$			
With Long-Run	'same-period'	0.9768	0.9921	1.0013	1.0025
		$\delta = 0$			
Without Long-Run	'same-period'	0.9781	0.9924	1.0014	1.0024

Table 6: RMSEs, Point-in-Time Sampling relative to MIDAS, correct number of variables

MIDAS vs Averaging		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'same-period'	1.1762	1.2008	1.2089	1.2109
		$\delta = -0.25$			
With Long-Run	'same-period'	1.0006	1.0200	1.0265	1.0270
		$\delta = -0.1$			
With Long-Run	'same-period'	1.0291	1.0469	1.0552	1.0550
		$\delta = 0$			
Without Long-Run	'same-period'	1.0701	1.0856	1.0933	1.0933

Table 7: RMSEs, Average Sampling relative to MIDAS, correct number of variables

for average sampled regressors are not surprising.

Furthermore, MIDAS yields superior forecasts compared to the Point-in-Time sampling for large sample sizes. For small T , however, Point-in-Time sampling results in lower RMSEs. Note that Point-in-Time sampling takes the correct m^{th} high-frequency observation as the corresponding low-frequency one. Therefore, it also includes the short-run variable having most impact in our DGP since the latter implies that more recent observations have higher impact.

Note also that, although not reported, the forecast accuracy of all approaches is most of the times not significantly different from that of MIDAS, exception made of Average sampling for $\delta = -0.75, -0.25$. Based on this, one might hence question the usefulness of MIDAS. However, we should emphasize the fact that until now, we have always assumed correctly adapted short-run dynamics, i.e. the chosen number of short-run variables is consistent with the methodology of Section 2.

Consequently, as a further analysis of the effect of possible short run misspecification, we study the sensitivity of the forecasting performance of MIDAS and the unrestricted approaches to an incorrect number of terms entering the short-run dynamics. In particular, as explained in the previous section, we investigate the situation where only $\frac{1}{2}m$ and as many as $3m$ short-run variables are included in the models (as opposed to the theoretically correct number of m variables). Note that the timing is fixed at the correct 'same-period'-case for this exercise and we only display the results for $\delta = -0.75$. Table 8 illustrates the outcomes for the case with too few variables whereas Table 9 gives the results for the case including too many. Again, the

figures represent the (means of the) RMSEs of the respective method relative to the benchmark case, namely MIDAS with the EC term and the correct number of short-run variables, m .

Too few short-run variables		$\delta = -0.75$			
Method / T		50	100	250	500
MIDAS with Long-Run		1.096 (68.2%)	1.111 (76.3%)	1.1121 (77.6%)	1.1111 (75%)
MIDAS without Long-Run		2.7667 (100%)	2.7975 (100%)	2.8589 (100%)	2.851 (100%)
Unrestricted with Long-Run		1.1431 (80.5%)	1.1284 (80.7%)	1.1202 (79.3%)	1.1127 (75.4%)
Unrestricted without Long-Run		2.8747 (100%)	2.8715 (100%)	2.88 (99.9%)	2.8667 (99.8%)

Table 8: RMSEs of MIDAS and Unrestricted Short-Run with too few variables relative to MIDAS with Long-Run and the correct number of variables

Too many short-run variables		$\delta = -0.75$			
Method / T		50	100	250	500
MIDAS with Long-Run		1.0578 (8.5%)	1.0324 (6.6%)	1.0117 (4.1%)	1.0033 (6.1%)
MIDAS without Long-Run		2.5133 (99.6%)	2.4991 (99.6%)	2.5145 (99.9%)	2.4979 (94.7%)
Unrestricted with Long-Run		2.2043 (84.2%)	1.2423 (33.1%)	1.0777 (8.6%)	1.0322 (6.5%)
Unrestricted without Long-Run		3.4667 (99.6%)	2.0603 (99.6%)	1.8039 (99.8%)	1.7142 (99.8%)

Table 9: RMSEs of MIDAS and Unrestricted Short-Run with too many variables relative to MIDAS with Long-Run and the correct number of variables

As becomes clear from Table 8, including too few short-run terms leads to a serious dynamic misspecification (serial correlation) of the model and, therefore, worsens the forecasting performances of both, MIDAS and unrestricted approach. Nevertheless, as shown by Miller (2011a), nonlinear least squares is consistent and minimizes the mean-squared forecast error even in the presence of serial correlation. The RMSEs of the dynamically misspecified models in Table 8 increase simply because the minimum mean-squared forecast errors increase in the presence of serial correlation. Furthermore, the figures in Table 9 reveal that including too many variables has significant negative consequences for the unrestricted method when T is small. This reflects the fact that the parameter proliferation is exacerbated since, now, 36 short-run variables need to be estimated (as opposed to 9 in the case of MIDAS). Again, this effect decreases as T increases. Note that the penalty for omitting the error-correction term when $\delta \neq 0$ is greater than the penalty for including superfluous short-run terms.

Note that for the MIDAS regressions, three separate weight functions are calculated, one for each set of m short-run variables. Since the last $2m$ variables entering the short-run dynamics are (according to Section 2) redundant, the three single weight functions capture this "kink" better than a single (smoother) weight function. Figure 1 shows the three separate weight functions for $T = 100$ when too many variables are included.

Tables 8 and 9 as well as Figure 1 show that the dynamic misspecification implied by

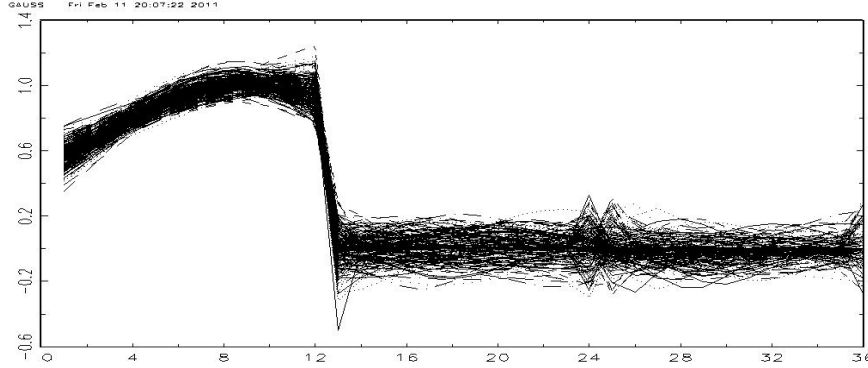


Figure 1: Weight functions, too many parameters, $T = 100$, Long-Run included

few terms in the short-run dynamics has a significant negative impact on the forecasting performance, irrespective of the approach employed. However, when the too many variables are included, MIDAS seems flexible enough to assign zero weight to the redundant variables whereas the unrestricted approach may suffer from parameter proliferation, the severity of which depending on the sample size. Hence, MIDAS' invariance to over-parameterization of the short run dynamics provides the user with a method that allows him to include enough short-run variables as necessary without knowing the true timing in the EC-term. The preserved parsimony of the model adds to this advantage of the MIDAS approach.

The previous analysis has an implication for (i) the selection of the timing prior to fixing the long-run relationship in the ECM and (ii) the methods to estimate the parameters. If one suspects a cointegrating relationship that is not "too strong" (e.g. with an expected small error correction coefficient), any of the candidate $x_{t,m-i}$ -terms may be used in the EC-term. It must be stressed that the number of short-run terms included in the ECM needs to be adapted to the respective timing chosen, unless one exploits the advantage of MIDAS in the presence of over-parameterized short run dynamics. Hence, the user may choose one of the candidate timings freely ('same period' for simplicity or 'x-after-y' due to data availability issues) as long as he adapts the short-run consistent with Section 2. Note that Miller (2011a) implicitly generates his cointegrating MIDAS regressions employing the 'same-period' case such that choosing this option would be in accordance with his work. MIDAS and its flexible data-driven weight functions provide an elegant way to ensure parsimony. If one suspects a "strong" cointegrating relationship, however, it appears that the timing may matter⁸ in which case we recommend applying a forecast combination method to determine weights for the candidate $x_{t,m-i}$ -terms. In particular, we advise to employ the approach proposed by Bates and Granger (1969) for its simplicity and consistency in putting most weight on the correct timing.⁹ As before, sufficient short-run dynamic terms need to be included in the ECM in order not to incur worsened

⁸See the case of 'x-after-y' with $\delta = -0.75$ when MIDAS is employed.

⁹The results of a corresponding simulation exercise may be provided on demand.

forecasting performances from including too few terms. Also in this situation, MIDAS provides a parsimonious choice that is invariant to the conservative approach of over-parameterization by including too many short-run variables.

4.4 High-Frequency DGP

So far, all simulations were done under the assumption that the DGP is of a mixed-frequency type. In order to investigate whether our results are invariant to another form of DGP, we consider a Gaussian VAR(p_h) $z_t^h = \sum_{i=1}^{p_l} \Phi_i z_{t-i}^h + \epsilon_t^h$ for the vector time series $\{z_t^h = (y_t^h : x_t^h)', t = 1, \dots, m, \dots, mT\}$. The error term ϵ_t^h , is an iid bivariate Gaussian vector with nonsingular covariance matrix Ω . Let us denote $\Phi(L) = I - \sum_{i=1}^{p_l} \Phi_i L^i$. Further assume that exactly one unit root is present so that $\text{rank}(\Phi(1)) = 1$ such that $\Phi(1)$ can be expressed as $\Phi(1) = -\alpha\beta'$ with α and β being (2×1) vectors of rank 1, where β is the cointegrating vector and α the error correction term. This can be rewritten in the desired error correction format:

$$\Delta z_t^h = \alpha\beta' z_{t-1}^h + \sum_{j=1}^{p_l-1} \Phi_j^* \Delta z_{t-j}^h + \epsilon_t^h, \quad (11)$$

with $\Phi_0^* = I$ and $\Phi_j^* = -\sum_{k=j+1}^{p_l} \Phi_k$ ($j = 1, \dots, p_l - 1$).

After y_t^h and x_t^h have been generated from this high-frequency VECM, Point-in-Time sampling on the last (end-of-period) observation or Average sampling is applied to y_t^h in order to convert it to low-frequency. In the sequel, the analysis follows the previous section. To this end the elements of the bottom rows of the matrices Φ_j^* in (11) are set to 0 to generate x_t^h as a random walk. The entries of the top row, i.e. the coefficients of the lagged terms of y_t^h and x_t^h in the equation for y_t^h are chosen to behave similarly to those in the previous section. In other words, the coefficients are increasing, but with decreasing step size such that more recent observations have a stronger impact than more distant ones.

As discussed in Miller (2011b), with white noise errors, a low-frequency model is consistent with the high-frequency model at hand only if the same aggregation scheme is chosen for both, the dependent and independent variables. Hence, Point-in-Time or Average sampling the dependent variable and choosing a different aggregation scheme for the regressors may introduce an estimation error (ϕ in Miller (2011b)) that enters the error term. On the contrary, the mixed-frequency DGP considered before assumed that sampling of the regressand does not affect the error term.

Table 10 presents the outcomes for a Point-in-Time sampled regressand where the short-run variables are modeled by MIDAS. The remaining results are similar to the outcomes of the previous section such that they are not reported for space reasons.¹⁰

In summary, as far as the individual forecasting methods are concerned, the main conclusions do not differ from those reached in the previous analysis. In other words, in the presence of cointegration, the inclusion of the error correction term improves the forecasting performances. Also, the forecasting performance of models with an erroneously chosen timing (but

¹⁰They are, however, available on request.

MIDAS		$\delta = -0.75$			
	Timing / T	50	100	250	500
With Long-Run	'x-before-y'	1.0022 (11.1%)	1.0026 (8.9%)	1.0021 (8.2%)	1.0016 (6.7%)
	x-after-y'	1.0984 (29)	1.0943 (28.3%)	1.0988 (29.6%)	1.0966 (28.3%)
Without Long-Run	'same-period'	1.3255 (99.6%)	1.3370 (99.8%)	1.3426 (99.6%)	1.3372 (99.6%)
		$\delta = -0.25$			
With Long-Run	'x-before-y'	0.9657 (16.7%)	0.9654 (14.7%)	0.9690 (13%)	0.9665 (14.6%)
	x-after-y'	1.0733 (23.8%)	1.0638 (21.7%)	1.0637 (20.7%)	1.0709 (25%)
Without Long-Run	'same-period'	1.2993 (99.7%)	1.3239 (99.8%)	1.3328 (100%)	1.3171 (100%)
		$\delta = -0.1$			
With Long-Run	'x-before-y'	0.9355 (23.8%)	0.9324 (22.5%)	0.9348 (18.7%)	0.9333 (20.7%)
	x-after-y'	1.0628 (23.3%)	1.0453 (17.4%)	1.0265 (12.9%)	1.0224 (12.7%)
Without Long-Run	'same-period'	1.2980 (99.8%)	1.3356 (99.9%)	1.3661 (100%)	1.3716 (99.9%)
		$\delta = 0$			
With Long-Run	'x-before-y'	0.9717 (37.2%)	0.9997 (32.6%)	1.0571 (36.9%)	1.1408 (45.9%)
	x-after-y'	1.0047 (40.1%)	1.0003 (25.9%)	1.0002 (15.7%)	1.0008 (11.8%)
Without Long-Run	'same-period'	0.9282 (28.4%)	0.9684 (17.1%)	0.9813 (8.2%)	0.9839 (5.2%)

Table 10: RMSEs relative to the 'same-period'-case with long run, the benchmark, and percentage number of rejection of tests of equal forecast accuracy at the 5% level compared to the benchmark, y by Point-in-Time sampling, short-run dynamics modeled by MIDAS, High-Frequency-DGP

with adequately adapted short-run dynamics) hardly differ from the benchmark case. Finally, in the absence of cointegration, correctly neglecting the EC term always results in the best outcome. Hence, the results of the previous section remain valid when the data are generated by a high-frequency DGP.

Similarly, the comparison of MIDAS with other approaches hardly changes. When the regressand is obtained by applying Point-in-Time sampling on y_t^h , MIDAS generally outperforms both the cases where the short-run is unrestricted and the case of Average sampling which performs the worst. Point-in-Time sampling performs as well as MIDAS which is not surprising since the dependent variable itself is Point-in-Time sampled. Indeed, Miller (2011b) and Chambers (2003) argue that, with white noise errors, the cointegrating vector is most efficiently estimated if the regressand aggregation scheme is matched by the one applied to the regressors. Note that Average sampling performs worst because it is not matching the regressand aggregation scheme. MIDAS is still able to adapt to the Point-in-Time sampled dependent variable due to its flexible data-driven weight determination whereas Average sampling imposes an erroneous flat weighting scheme.

If y_t^h is temporally aggregated by Average sampling, the results are again essentially similar, with the obvious exception that the roles of Point-in-Time and Average sampled regressors are

interchanged. Hence, if the aggregation scheme underlying the dependent variable is known, it seems best to apply the same aggregation method to the regressors. Nevertheless, applying MIDAS yields equally good forecasts. If, however, the regressand aggregation scheme is unknown, applying an erroneous classical aggregation method may deteriorate the forecasting performance immensely whereas MIDAS is able to mimic the respective aggregation scheme quite well and therefore provides a robust, yet parsimonious solution to the problem.

5 Applications

As an illustration, we consider a small model to forecast US inflation. All the variables have been downloaded from the Federal Reserve Bank of St. Louis at <http://research.stlouisfed.org/fred2/> and cover the period 1980 – 2010. The dependent variable denoted CPI_t is the seasonally adjusted US monthly consumer price index for all urban consumers (all items). We have collected weekly indicators but only report the results for the model with the best fit. The explanatory variables are (i) the US regular all formulations gasoline price (GAS_t) and (ii) the S&P500 index (denoted SP_t). Interest rates of different maturities, the weighted exchange index over major currencies as well as the stock of money M2 did not improve over the regression presented here. We estimate the model on the period ranging from January 2000 to November 2010, amounting to 131 observations. We have focused on that period in order to have a price index series that is well described as $I(1)$ and, hence, to avoid the discussion about the potential $I(2)$ -ness of that series over a longer span.

In contrast to the previous section, the cointegrating vector is assumed unknown and hence estimated. Also, note that Miller (2011a)’s asymptotic results are of importance here. It is likely that serial correlation is an issue in this application, yet the consistency of the nonlinear least squares estimator to the minimum mean-squared forecast error parameter vector remains valid as well as the asymptotic distribution of the difference between the two. In other words, even if serial correlation is present here, nonlinear least squares consistently estimate the minimum mean-squared forecast error parameter vector (Miller, 2011a).

As far as the presence of a long-run relationship between $\ln(CPI_t)$ and the two regressors in a multivariate static regression is concerned, we consider a window of three weeks before and after the contemporaneous relationship. We thus estimate static cointegrating regressions on $x_{t,m-i}$ with $i \in [-3, 3]$ because $m = 4$ in a month/week setting. We further assume that the explanatory variables enter the regression with the same time shift. This restriction may easily be relaxed but it is imposed here to save on computation burden. The results of the ADF and Johansen’s Trace tests (Johansen, 1991) are reported in Table 11 below.¹¹

The ADF and Johansen’s trace tests reject the null of no-cointegration for all candidate timings. This should come as no surprise given that cointegration is a long-run property that is invariant to temporal aggregation Marcellino (1999). In other words, if cointegration exists for

¹¹Johansen’s Trace Test is computed while allowing for a linear deterministic trend in the data and an intercept, but no trend in the cointegrating equation(s). Furthermore, the optimal lag length in a VAR is 2 based on SIC such that 1 lag of differences is chosen for a VECM.

Error-Correction term	Timing	ADF(1)	Johansen's Trace Test for no CE(s) p-value	Johansen's Trace Test for at most 1 CE p-value
$y_{t-1} - \beta x_{t,3}$	'x-after-y'	-4.5917	0.0179	0.4967
$y_{t-1} - \beta x_{t,2}$		-4.674	0.0083	0.3874
$y_{t-1} - \beta x_{t,1}$		-4.7754	0.0039	0.4010
$y_{t-1} - \beta x_{t-1,4}$	'Same Period'	-4.6107	0.0101	0.4755
$y_{t-1} - \beta x_{t-1,3}$	'x-before-y'	-4.4277	0.0199	0.5114
$y_{t-1} - \beta x_{t-1,2}$		-4.7	0.0047	0.5323
$y_{t-1} - \beta x_{t-1,1}$		-4.8408	0.004	0.6115

Table 11: ADF test statistics and Johansen's Trace test for the application

one combination of y_t and $x_{t,m-i}$, it should do so for the other timings of the regressors as well. With respect to the regressor variable entering the long-run we choose the 'same-period'-case due to its simplicity and accordance with Miller (2011a). Estimating the long-run term by fully modified least squares yields (where lower case letters represent logarithms of the respective variables)

$$\widehat{z_{t-1}} = \widehat{ecm_{t-1}} = cpi_{t-1} - \underset{(0.201685)}{6.274969} + \underset{(0.029048)}{0.170431}sp_{t-1} - \underset{(0.015342)}{0.275665}gas_{t-1}.$$

Every coefficient is individually significantly different from zero at any sensible significance level (FMOLS standard errors in brackets). Plugging this into an error-correction model we can compare the forecasts of different modelings for the last 23 months in Table 12. Note that $p_l = 2$ and $q_l = 1$ such that there is one lagged difference of cpi_t as well as 8 short-run terms per regressor in the ECM. Note that several weight functions per low-frequency period were estimated for the MIDAS regressions, i.e. for every regressor one weight function was estimated per low-frequency period considered (4 weight functions in total).

Regressor Aggregation Scheme	RMSE
Point-in-Time sampling with Long-Run	0.00968
Point-in-Time sampling without Long-Run	0.01065
Average sampling with Long-Run	0.00837
Average sampling without Long-Run	0.00954
Unrestricted with Long-Run	0.00921
Unrestricted without Long-Run	0.01071
MIDAS with Long-Run	0.0088
MIDAS without Long-Run	0.01012

Table 12: RMSEs for different methods to forecast monthly US inflation using weekly indicators

As expected, the RMSEs of the models with an EC term are lower than those without. However, the differences are quite small. Closer investigation of the respective estimates of

error correction coefficients ($\hat{\delta}$) reveals that the speed of adjustment is very slow relative to the simulations conducted before. Furthermore, in the simulations, the coefficients as well as the values of x and y themselves were much larger than in the application which yielded larger RMSEs and, thereby, larger differences between them. To formally test whether the RMSEs of including and excluding an ECM term differ from each other, the modified Diebold-Mariano test of Harvey et al. (1998) is conducted for each method. The statistics for Point-in-Time and Average sampling, the unrestricted approach and MIDAS are 3.186, 3.32, 4.544 and 2.851, respectively. Hence, for all methods under consideration, including a long-run term leads to significantly lower RMSEs and, thereby, better forecasting performances.

Referring again to Table 12, Average sampling performs best among the four methods applied here. Relying on (modified) Diebold-Mariano tests reveals that, when including an ECM term, the forecast accuracies of the four methods are not significantly different from each other, with the exception being Point-in-Time sampling which yields a significantly larger RMSE. Foroni et al. (2011) investigate the difference between so-called unrestricted MIDAS (U-MIDAS) and traditional MIDAS models. For different frequency combinations, i.e. different m in our notation, they compare the forecasting (or nowcasting) accuracy of a model where the high-frequency variables are estimated by least squares (unrestrictedly) with a model where a nonlinear MIDAS structure is imposed and the variables are estimated by nonlinear least squares. They find that for smaller m , e.g. $m = 3$, the unrestricted method dominates traditional MIDAS models whereas for larger m , e.g. $m = 12$ or especially $m = 60$, MIDAS clearly outperforms the unrestricted approach. This could explain why MIDAS does not significantly outperform the unrestricted method here.

Note finally that the estimation period still consists of 109 observations which seems still too much for the unrestricted approach to run into parameter proliferation issues. Although MIDAS is not significantly outperforming the other methods, it easily enables us to extend the above analysis. Adding more high-frequency regressors and/or more lags comes at no cost in terms of parsimony whereas the unrestricted method would quickly suffer from the aforementioned problem of parameter proliferation while Average sampling would ignore high-frequency-information present in the new variables. Hence, MIDAS remains a robust and parsimonious approach in such real-life situations.

6 Conclusion

In this paper, a mixed frequency ECM is proposed in order to model non-stationary variables that are possibly cointegrated. It is shown that, in terms of timing, the choice of variables entering the error-correction term has an impact on the short-run dynamics of the mixed-frequency ECM at the model representation level. In particular, three cases have been distinguished, the 'same-period', the 'x-before-y'- and the 'x-after-y'-case and the corresponding error correction specifications have been developed.

We propose to use a simple approach inspired by the Engle-Granger two-step framework (Engle and Granger, 1987) in order to decide on the composition of the long-run relationship.

In particular, one may either choose one of the regressor variables freely or apply a forecast combination method depending on the application at hand. An alternative mixed-frequency approach is discussed as well and its disadvantages to the previous MF-ECM approach are elaborated on.

The forecasting performances of four approaches (unrestricted short-run, MIDAS regression, Point-in-Time and Average sampling) with and without the inclusion of an long-run term are assessed in Monte Carlo simulations. We also consider various possible problems such as using a "mis-timed" long-run relationship as well as over- and under-parameterization of the short-run dynamics. To assess the results of considering an erroneous timing in the EC term, including $x_{t-1,m-6}$ ('x-before-y') and $x_{t,m-6}$ ('x-after-y') instead of $x_{t-1,m}$ ('same-period') into the long-run term is considered. For the latter problem we consider both the cases where too few and too many short-run variables are included in the models.

Several important conclusions were drawn on the basis of these simulations. Firstly, ignoring the EC term significantly lowers the forecasting performance if the variables are cointegrated. Thus, if a practitioner deals with non-stationary variables, it should be checked whether the variables are cointegrated prior to transforming them into stationary variables. This is in line with the results of Clements and Hendry (1998). If they are cointegrated, the inclusion of an EC term will significantly improve his forecasts. Secondly, in almost all cases, a "mis-timed" cointegrating relationship, though leading to a slightly larger RMSEs, does not result in a significantly lower forecast accuracy than the correct model provided the short-run dynamics are adapted accordingly. Hence, an erroneous timing does not introduce a misspecification to the model, but rather introduces an alternative representation. As a consequence, the regressor variable entering the EC-term may be chosen freely and we advise the 'same-period'-case due its simplicity and agreement with Miller (2011a).

Thirdly, under-parameterization of the short-run dynamics on the one hand leads to a misspecification of the model resulting in negative impacts on the performance of all approaches. Over-parameterization on the other hand does not lead to worsened forecasting performances although the increased number of coefficients to estimate may deteriorate the outcomes of the unrestricted approach for small sample sizes. MIDAS does not suffer from this problem due to its ability to include many high-frequency variables in a parsimonious way. It therefore provides a robust alternative to include as many short-run variables as necessary given the underlying DGP and yet retain a parsimonious model.

Finally, MIDAS outperforms the unrestricted approach for small sample sizes emphasizing its advantage in terms of parameter proliferation. The forecast accuracy of MIDAS and the two classical aggregation methods as well as the unrestricted method for large sample sizes do not differ significantly from each other. The main findings just described seem to be invariant to an alternative DGP where both, regressand and regressor, are generated at the high frequency.

A small empirical application is used to illustrate the discussion of Section 2 as well as the Monte Carlo results. In particular, US monthly inflation is forecasted employing several weekly variables. It is found that including the disequilibrium error, i.e. the EC term, significantly enhances the forecasting performance. Furthermore, Average sampling yields the lowest

RMSE suggesting a flat underlying regressand aggregation scheme which would explain why Point-in-Time sampling generates the largest RMSE. MIDAS, Average sampling and modeling the short-run unrestrictedly yield no significantly different forecast accuracies from each other although MIDAS allows to straightforwardly extend the analysis if desired whereas the other two methods would either suffer from the curse of dimensionality (unrestricted method) or disregard information present in eventually added high-frequency regressors (Average sampling). Overall, MIDAS proves to be a robust choice in case the true regressand aggregation scheme is unknown or in case the sample size is not large enough.

References

- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):pp. 178–196.
- Andreou, E., Ghysels, E., and Kourtellis, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.
- Andreou, E., G. E. and Kourtellis, A. (2010). Should macroeconomic forecasters use daily financial data and how? Working paper series, Rimini Centre for Economic Analysis.
- Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *OR*, 20(4):pp. 451–468.
- Chambers, M. J. (2003). The asymptotic efficiency of cointegration estimators under temporal aggregation. *Econometric Theory*, 19(1):pp. 49–77.
- Clark, T. E. and McCracken, M. W. (1999). Tests of equal forecast accuracy and encompassing for nested models. Research Working Paper 99-11, Federal Reserve Bank of Kansas City.
- Clements, M. and Galvao, A. B. (2007). Macroeconomic forecasting with mixed frequency data: Forecasting us output growth. Working Papers 616, Queen Mary, University of London, School of Economics and Finance.
- Clements, M. and Galvao, A. B. (2009). Forecasting us output growth using leading indicators: an appraisal using midas models. *Journal of Applied Econometrics*, 24(7):1187–1206.
- Clements, M. and Hendry, D. (1998). *Forecasting Economic Time Series*. Number 9780521634809 in Cambridge Books. Cambridge University Press.
- Davidson, J. (2000). *Econometric Theory*. Blackwell Publishers Ltd.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):pp. 253–263.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–76.
- Engle, R. F. and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35(1):143–159.
- Forni, C., Marcellino, M., and Schumacher, C. (2011). U-midas: Midas regressions with unrestricted lag polynomials. Research working paper.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.

- Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.
- Ghysels, E. and Valkanov, R. (2006). Linear time series processes with mixed data sampling and midas regression models. Discussion Paper.
- Götz, T. B. (2010). Mixed data sampling for nonstationary time series. Master’s thesis, Maastricht University, School of Business and Economics.
- Harvey, D., Leybourne, S., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, 16(2):254–59.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):pp. 1551–1580.
- Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–36.
- Miller, J. I. (2011a). Cointegrating midas regressions and a midas test. Working Papers 1104, Department of Economics, University of Missouri.
- Miller, J. I. (2011b). Conditionally efficient estimation of long-run relationships using mixed-frequency time series. Working Papers 1103, Department of Economics, University of Missouri.
- Pons, G. and Sansó, A. (2005). Estimation of cointegrating vectors with time series measured at different periodicity. *Econometric Theory*, 21(04):735–756.
- Seong, B., Ahn, S. K., and Zdrozny, P. (2007). Cointegration analysis with mixed-frequency data. CESifo Working Paper Series 1939, CESifo Group Munich.
- Silvestrini, A. and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. Temi di discussione (Economic working papers) 685, Bank of Italy, Economic Research Department.